In: European Journal of Psychology of Education, 2016, Volume 31, Issue 2, S. 117-134, EISSN 1878-5174; https://doi.org/10.1007/s10212-015-0251-z, Accepted Manuscript

Author Information

Children's Use of Number Line Estimation Strategies

Dominique Peeters[a], Tine Degrande[a], Koen Luwel[ab], Mirjam Ebersbach[ac], and Lieven Verschaffel[a]

[a] Centre for Instructional Psychology and Technology, KU Leuven, Belgium

[b] Centre for Educational Research and Development, KU Leuven – Campus Brussels, Belgium

[c] Institut für Psychologie, Universität Kassel, Germany

Author Note

Dominique Peeters, Centre for Instructional Psychology and Technology, KU Leuven; Tine Degrande, Centre for Instructional Psychology and Technology, KU Leuven; Koen Luwel, Centre for Educational Research and Development, KU Leuven – Campus Brussels and Centre for Instructional Psychology and Technology, KU Leuven; Mirjam Ebersbach, Institut für Psychologie, Universität Kassel; Lieven Verschaffel, Centre for Instructional Psychology and Technology, KU Leuven.

The conduct of this study was supported by grant GOA 2012/010 of the Research Fund KU Leuven, Belgium and by grant DFG: EB462/1-1 of the German Research Foundation to the fourth author. Dominique Peeters and Tine Degrande both contributed equally to the study.

Correspondence concerning this article should be addressed to Dominique Peeters, Centre for Instructional Psychology and Technology, KU Leuven, Dekenstraat 2, box 3773, B-3000 Leuven, Belgium.

Telephone Number: +32 16 37 31 65

E-mail: dominique.peeters@ppw.kuleuven.be

Abstract

This study tested whether second graders use benchmark-based strategies when solving a number line estimation (NLE) task. Participants were assigned to one of three conditions based on the availability of benchmarks provided on the number line. In the bounded condition, number lines were only bounded at both sides by 0 and 200, while the midpoint condition included an additional benchmark at the midpoint and children in the quartile condition were provided with a benchmark at every quartile. First, the inclusion of a midpoint resulted in more accurate estimates around the middle of the number line in the midpoint condition compared to the bounded and, surprisingly, also the quartile condition. Furthermore, the two additional benchmarks in the quartile condition did not yield better estimations around the first and third quartile, because children frequently relied on an erroneous representation of these benchmarks, leading to systematic estimation errors. Second, verbal strategy reports revealed that children in the midpoint condition relied more frequently on the benchmark at the midpoint of the number line compared to the bounded condition, conforming the accuracy data. Finally, the frequency of use of benchmark-based strategies correlated positively with mathematics achievement and tended to correlate positively also with estimation accuracy. In sum, this study is one of the first to provide systematic evidence for children's use of benchmark-based estimation strategies in NLE with natural numbers and its relationship with children's NLE performance.

Key words: Number line estimation, benchmarks, estimation strategies

Numbers play a prominent role in our daily life. They are used to specify quantities, time, distances, wealth, risks, and many other quantifiable features of objects, people, and situations. Recent studies have shown that the mental representation of numerical magnitudes is related to and predictive for children's mathematics achievement (e.g., Booth & Siegler, 2006; Bugden & Ansari, 2011; De Smedt, Verschaffel, & Ghesquière, 2009; Gilmore, McCarthy, & Spelke, 2010; Sasanguie, De Smedt, Defever, & Reynvoet, 2012). These findings stress the importance of the ability to understand and process numerical information for success in education and life (Ancker & Kaufman, 2007; Finnie & Meng, 2001).

During the last decades, the number line estimation (NLE) task became a widely used tool for investigating an individual's mental magnitude representation. In this task, participants are asked to either estimate the spatial position of a number on an empty, physical number line with labeled endpoints (e.g., 0 and 100 or 1000), or, alternatively, estimate the number which corresponds to a given spatial position on such number line. These tasks are known as the number-to-position (NP-task) and position-to-number task (PN-task), respectively. The pattern that emerges when plotting the estimated positions of the numbers on the number line as a function of their actual position, has been assumed to reflect an individual's mental representation of that particular number range (Siegler & Opfer, 2003).

Siegler and Opfer (2003) investigated the development of numerical magnitude representations by having a group of second, fourth, and sixth graders as well as adults perform the NLE task on a 0 □ 100 and 0 □ 1000 scale. They demonstrated that, with increasing age, participants' NLE patterns evolved from a logarithmic towards a linear pattern on the 0 □ 1000 scale, whereas all age groups exhibited a linear pattern on the 0 □ 100 scale. This finding suggests that, with increasing age and experience with a specific number range, children's underlying magnitude representation develops from a logarithmically compressed

mental number line with increasing numerical magnitudes being successively more closely spaced (Dehaene, 1997) towards a linear representation, reflecting the equal spacing principle of the mature number system. Notably, however, some authors have argued that young children's estimation patterns might be described better by a two-segmented linear model than by a logarithmic one (Ebersbach, Luwel, Frick, Onghena, & Verschaffel, 2008). The steep slope of the first segment refers to the range of numbers children are familiar with and which they can differentiate easily. Beyond that familiar number range, discrimination between numbers is more difficult, resulting in a second segment with a rather shallow slope. Hence, the change point between the two linear curves functions as an indicator of number familiarity (see also Moeller, Pixner, Kaufmann, & Nuerk, 2009, for a similar two-linear approach suggesting separate but linear representations for single- and two-digit numbers).

The so-called log-to-lin or representational shift in children's estimation patterns has been replicated at different ages and for different scales (Berteletti, Lucangeli, Piazza, Dehaene, & Zorzi, 2010; Siegler & Booth, 2004; Thompson & Opfer, 2010; see Siegler, Thompson, & Opfer, 2009 for a review) as well as in other types of estimation tasks, such as estimating quantities or measurements (Booth & Siegler, 2006). Furthermore, the linearity of NLE patterns is strongly related to estimation accuracy (Ashcraft & Moore, 2012; Siegler & Booth, 2004), performance in basic numerical tasks (Berteletti et al., 2010), and general mathematics achievement as well as other measures of mathematical ability (Ashcraft & Moore, 2012; Booth & Siegler, 2006, 2008; Schneider, Grabner, & Paetsch, 2009; Siegler & Booth, 2004).

Notwithstanding these robust findings, several sources of evidence question whether the NLE task can be considered as a pure measure of an individual's underlying numerical magnitude representation and propose that improvements in NLE performance might be explained at least to a significant part by an increasing reliance on a variety of

benchmark-based NLE strategies, rather than by changes in children's mental analog of the physical number line per se.

Firstly, by relying on verbal self-reports, Newman and Berger (1984) found that third graders, but not first graders, made use of the midpoint when estimating numbers in the middle range of the number line. Secondly, by closely observing children's overt solution behavior and thus which strategies they use, Petitto (1990) found that the percentage of children using a midpoint strategy increased across grades as arithmetic and counting skills improved. Thirdly, Ashcraft and Moore (2012) observed specific patterns in children's error rates and latencies that deviated strongly from the patterns expected if they would simply be accessing a mental analog of the physical number line (see also White & Szucs, 2012). More specifically, with age, error rates and latencies first started to decrease at the endpoint of the number line, and later at the midpoint as well, resulting in a typical M-shaped pattern, suggesting the use of an endpoint and a midpoint strategy, respectively. Fourth, Siegler and Opfer (2003) observed that adults' and sixth graders' estimates on a 0–1000 number line were less variable near 0, 250, 500, 750, and 1000 in comparison to other numbers, suggesting that participants divided the number line into quarters, which were then used as benchmarks to guide their estimates. Fifth, Barth and Paladino (2011) demonstrated that, with age, children's NLE patterns are successively best described by an unbounded power model, a one-cycle power model, and a two-cycle power model (see also Slusser, Santiago, & Barth, 2013). According to these authors, this development in model complexity reflects children's increasing sophistication in the use of benchmarks on the number line. That is, the unbounded power model characterizes NLE patterns for which only the origin is taken into account. The best fit with a one-cycle model reflects the strategic use of both origin and endpoint, while a best fit with a two-cycle model suggests the reliance on the midpoint too. In light of these results, Barth and Paladino (2011) suggested that the typical NLE task should be considered

as a proportion judgment task instead of a numerical magnitude estimation task. Finally, to verify whether the presence of an endpoint indeed leads to estimation patterns that are better fit by a one- or two-cycle power model instead of an unbounded power model, Link, Huber, Nuerk, and Moeller (2014) recently tested children from first to fourth grade on both an unbounded (i.e., without an endpoint) and bounded NLE task. Results indicated that, in the unbounded version of the task, children's estimates in all age groups were best fit by an unbounded power model, indicating that they solely relied on the origin of the number line to make their estimates. In the bounded number line task, however, children's estimation patterns were, in line with Barth and Paladino's (2011) findings, successively best fit by an unbounded, a one-cycle, and a two-cycle power model as age increased. Taken together, these studies indicated that as children develop, they start using external (i.e., the endpoint) or self-generated internal (e.g., the midpoint) benchmarks on the number line to guide their estimates.

Although it has become increasingly clear that children appear to use a variety of benchmark-based strategies when making NLEs, and that this variation in strategy use accounts for their NLE performance, the precise nature of these strategies has not yet been studied in a direct and systematic way. The goal of the present study was to provide such a direct and systematic analysis, by confronting children with number lines containing different numbers of external benchmarks and by asking them to verbally explain how they solved each NLE item, in addition to an analysis of their error rates. To the best of our knowledge, this is the first study to include such external benchmarks not only at the midpoint but also at the first and third quartile when estimating the position of natural numbers on a number line. Furthermore, while the technique of verbal protocols has already been intensively used in cognitive research on children's mathematical strategies in general (e.g., Siegler & Stern, 1998; Torbeyns, De Smedt, Ghesquière, & Verschaffel, 2009), and also in one older study on

children's NLE (Newman & Berger, 1984), the present study is the first wherein young children's verbal strategy reports have been used in a systematic way to help unravel whether and how they rely on externally provided and/or self-generated internal benchmarks on a number line.

So, in the present study, estimations had to be completed in one of three conditions that differed with respect to the availability of particular benchmarks on a 0 to 200 number line (see Figure 1). In the bounded condition, number lines were bounded at both sides by the corresponding benchmarks (i.e., 0 and 200). In the midpoint condition, bounded number lines with an additional benchmark at the midpoint (i.e., at 100) were presented. In the quartile condition, children were provided with a bounded number line with a benchmark at every quartile (i.e., at 50, 100 and 150). Three sets of research questions were addressed.

A first set of research questions addressed whether providing additional external benchmarks had a positive effect on children's NLE performance. As mentioned earlier, with increasing age, children have a greater tendency to use an internal benchmark at the midpoint when making NLEs, resulting in an improved NLE performance (e.g., Ashcraft & Moore, 2012, Barth & Paladino, 2011). Following Siegler and Opfer's (2003) finding that estimates on a 0–1000 number line were less variable near 0, 250, 500, 750, and 1000, which suggested the possible use of internal benchmarks at the quartiles too, we anticipated that providing extra benchmarks at the first and third quartile would further positively affect estimation accuracy. In sum, we expected that an increase in the number of given benchmarks would lead to higher overall estimation accuracy. Secondly, the provision of benchmarks was expected to affect the accuracy on the items located near the respective provided benchmarks. More specifically, we predicted that estimations around the first and third quartile would be more accurate in the quartile than in the midpoint and bounded condition, whereas the

estimations around the midpoint would be more accurate in the quartile and midpoint than in the bounded condition.

A second set of research questions focused specifically on children's strategy use. Firstly, we expected that an increase in the number of provided benchmarks would result in a greater frequency and variety of reported standard benchmark-based (SBB) strategies (i.e., strategies that made use of self-generated or provided benchmarks that were located at the quartiles of the number line). Secondly, regarding the frequency of these SBB strategies, we expected that children would make more frequently use of strategies that were based on the benchmarks provided in the respective conditions. More specifically, we predicted a relatively high frequency of strategies based on the origin and endpoint of the number line in the bounded condition, whereas in the midpoint condition we expected more strategies based on the midpoint too, while in the quartile condition strategies based on the given benchmarks at all quartiles were expected. Thirdly, if children do make use of the benchmarks provided to them, the use of SBB strategies should be reflected in their estimation patterns. Based on previous research (Barth & Paladino, 2011; Slusser et al., 2013), we hypothesized that children's estimations in the bounded condition would be best fit by a one-cycle power model, reflecting the strategic use of both the origin and endpoint when making NLEs. Children in the midpoint condition were expected to be best fit by the two-cycle power model since they can rely on the midpoint too, whereas children in the quartile condition should be best fit by a four-cycle power model indicating the use of strategies based on the origin, endpoint, midpoint, first, and third quartile.

As a third and final research question, we investigated the extent to which the use of SBB strategies would be related to children's NLE accuracy, as well as their general mathematics achievement.

Method

Participants

Sixty-four second graders (34 boys, 30 girls, *M* = 8.06 yrs., *SD* = 0.42 yrs.) were recruited from four classes in two elementary schools located in a rural area of Flanders (Belgium). Most children came from middle-income families. In none of the classes NLE had been systematically taught. All children participated voluntarily with informed consent of their parents and teachers. Children were told that they could quit the experiment at any moment.

Materials

Children's NLE performance was assessed by a NP-task with a number line labeled at the left end by "0" and at the right end by "200". According to the elementary school curriculum and the arithmetic textbook used, second graders had only systematically explored the number range up to 100 at the moment of the data collection (February/March), while the 100 □ 200 range had received no instructional attention at all. To avoid ceiling effects in NLE accuracy and shape of estimation patterns when using a 0-100 number line in second graders (cf. Siegler & Opfer, 2003), a 0-200 number line was used.

For each trial, a new number line with a length of 25 cm was presented on a separate sheet of paper. To avoid that the presented number might function as an additional benchmark, it was positioned completely on the left of the page. Children had to position 20 randomly chosen numbers on the number line, one from each decade between 0 and 200. These 20 numbers were equally distributed across the number line in order to prevent overestimation of smaller numbers due to oversampling at this end of the number line (Ebersbach et al., 2008). The 20 to-be-positioned numbers were: 5, 16, 22, 38, 43, 55, 62, 76, 87, 91, 103, 110, 129, 134, 146, 159, 162, 173, 189, and 194. The presentation order of these 20 numbers was randomized across participants. Children also received six practice trials.

Numbers used in the practice trials were also equally spread over the 0 to 200 number range: 4, 14, 46, 105, 141, 168. When the estimate of the first or second practice trial was very inaccurate, instructions were shortly repeated. However, at no point individual feedback was provided.

Children were randomly assigned to one of three conditions (see Figure 1). Special care was taken to ensure that the number of boys and girls was about equal in the three conditions. In the bounded condition, an empty number line on which the origin and endpoint were indicated by a small vertical line at either side of the number line was presented. The corresponding numbers, 0 and 200, were printed underneath these vertical lines. In the midpoint condition, a benchmark at the middle of the number line was included by presenting a vertical line at the position of 100 but no number. Finally, in the quartile condition, three extra benchmarks were given by introducing vertical lines at positions 50, 100, and 150 of the number line. Thus, the benchmarks at 50, 100, and 150 in the midpoint and quartile condition, were only represented by a vertical line without its corresponding number.

To conclude, children's mathematics achievement was measured by means of the standardized mathematics test of the Flemish Student Monitoring System (Dudal, 2000) for the middle of the second grade.

(Figure 1 about here)

Procedure

The NLE task was administered individually in a quiet room at the school. Children were given following instructions: "I'm going to show you some number lines. These number lines start at 0 and end at 200. For each trial, a number between 0 and 200 is shown on the upper left side of the page. What I want you to do is to put a mark on the line where you think the number would go." The to-be-positioned numbers were read out loud by the experimenter

to ensure that children knew which number had to be placed on the number line. Immediately after each trial, children reported which strategy they had used. In case the child produced an unclear verbal report, the experimenter asked non-intrusive follow-up questions such as "How did you do that?" and "What were you thinking?". At the end of the task, children in the midpoint and quartile condition were asked to indicate which number corresponded to the vertical line at 100 (midpoint condition) or 50, 100, and 150 (quartile condition). The NLE task lasted 30 to 60 minutes per child and was completely recorded with a voice recorder.

Data Analyses

To determine the accuracy of number line estimates, positions of children's handwritten marks on the number lines were measured. The measured distances were converted into numerical estimates. For these numerical estimates, we then calculated the percentage absolute error (PAE) by means of the following formula (Siegler & Booth, 2004):

$$\frac{|estimate - estimate\ quantity|}{scale\ of\ estimates} \times 100$$

Strategy reports were analyzed through a self-designed classification scheme consisting of three main categories: perceptual strategies, SBB strategies, and non-SBB strategies (see Figure 2). Perceptual strategies include seeing, guessing, and knowing (Gandini, Ardiale, & Lemaire, 2010; Gandini, Lemaire, & Dufau, 2008). SBB and non-SBB strategies refer to strategies that make use of a benchmark. Specifically, SBB strategies make use of self-generated or externally provided benchmarks that are located at the quartiles on the number line: origin (0%), first quartile (25%), midpoint (50%), third quartile (75%), and endpoint (100%). For instance, when a child had to locate 44 on the 0-200 number line and reported that (s)he had first looked at the benchmark representing 50 and then located 44 a

little bit to the left of that benchmark because 44 is 6 less than 50, it was coded as 25%[1]. Non-SBB strategies are based on other, self-generated benchmarks different from those five standard benchmarks, such as a benchmark at 33% or 12.5% of the number line or a benchmark corresponding to the child's answer to a previous item. Finally, a rest category was available for cases that were not classifiable in one of the above-mentioned categories.

Reliability of the classification scheme was assessed by testing the agreement in classification of two independent raters who classified all experimental trials of eight randomly chosen participants by means of Cohen's Kappa. This inter-rater reliability measure was .89, indicating almost perfect agreement according to the standards of Landis and Koch (1977).

According to Ericsson and Simon (1984; see also McGilly & Siegler, 1989), verbal reports tend to be most valid when they are immediately retrospective, refer to an individual's global solution process rather than its finest details, and the strategy under investigation takes several seconds or even more to execute. Given that the participants in the present study had an average latency per estimation of 9 s (*SD* = 3 s) in each of the three conditions, and provided their retrospective reports immediately after solving each item, we were quite confident about the validity of these reports.

(Figure 2 about here)

[1] Note that such a verbal report was coded as 25% even when the child actually pointed to the 50% benchmark when saying "first I looked at the benchmark representing 50". In other words, when relying on the verbal protocols we coded the benchmarks as represented by the child even when they had misrepresented it (e.g., when they erroneously thought that the 25% benchmark represented 100 instead of 50).

Results

Estimation Accuracy

For each of the three conditions, individual estimates that deviated more than 2 *SD* from children's mean estimate for the to-be-positioned number were excluded. In total, 71 estimates out of a total of 1280 were removed (5.5%). These 71 estimates were similarly distributed across conditions (bounded: 24, midpoint: 29, and quartile: 18). Tukey HSD tests were used in all post hoc comparisons.

Overall estimation accuracy. A one-way ANOVA assessing the effect of condition (bounded, midpoint, quartile) on overall PAE, was significant, $F(2, 61) = 16.66$, $p < .0001$. Estimates in the quartile condition ($M = 14.88$, $SD = 5.26$) were, contrary to our predictions, significantly less accurate than in the midpoint ($M = 7.32$, $SD = 4.39$, $p < .0001$) and even the bounded condition ($M = 9.66$, $SD = 5.26$, $p = .0007$), while there was no significant difference between the latter two conditions. So, the provision of additional benchmarks in the quartile condition seemed to be counterproductive instead of helpful for the overall accuracy of children's NLEs.

Estimation accuracy near the benchmarks. To get a closer view on the accuracy near the location of the standard benchmarks, we conducted a contour analysis on the trials where SBB strategies were used (Ashcraft & Moore, 2012). Therefore, we averaged on a child-by-child basis the observed PAEs for the two items located immediately before and after the midpoint (i.e., 91 and 103) and quartiles (i.e., 43 and 55; 146 and 159). For the origin and endpoint, our measure was based on the PAE of the number immediately after the origin and before the endpoint, respectively (i.e., 5 and 194). A 3 (Condition: bounded, midpoint, quartile) × 5 (Location: 0%, 25%, 50%, 75%, 100%) ANOVA with repeated measures on the last variable and the PAEs as dependent variable revealed a significant main effect of

condition ($F(2, 61) = 19.47$, $p < .0001$). Estimates in the midpoint condition ($M = 5.04$, $SD = 2.50$) were significantly more accurate compared to the bounded ($M = 8.89$, $SD = 5.41$, $p < .0001$) and the quartile condition ($M = 9.91$, $SD = 7.54$, $p < .0001$). No difference was found between the latter two conditions.

Furthermore, a significant interaction effect between condition and location was observed ($F(8, 244) = 5.32$, $p < .0001$) (see Figure 3). No significant difference between conditions at the origin, third quartile, and endpoint was found. However, at the first quartile, estimates in the quartile condition ($M = 15.35$, $SD = 9.18$) were significantly less accurate than estimates in the midpoint condition ($M = 6.80$, $SD = 5.32$, $p < .0001$), which is opposite to our expectation. At the midpoint, estimates in the midpoint condition ($M = 6.95$, $SD = 8.60$) were significantly more accurate than in the bounded ($M = 14.89$, $SD = 8.33$, $p = .0004$) and even the quartile condition ($M = 19.19$, $SD = 9.00$, $p < .0001$). Thus, the inclusion of a midpoint did result in more accurate estimates around the middle of the number line in the midpoint condition but not in the quartile condition. Furthermore, the two additional benchmarks in the quartile condition did not lead to better estimates around first and third quartile in that condition.

(Figure 3 about here)

Nature of the estimation errors. In search of an explanation for the remarkable finding that estimation performance around the benchmarks was not better in the quartile condition compared to the bounded condition, we analyzed the estimation data of this condition in greater detail.

First, we looked at the items at 25%, 50%, and 75%[2] of the number line (i.e., 43, 55, 91, 103, 146, 159). A closer inspection of children's estimated positions revealed that children from the quartile condition gave estimates that deviated *systematically* from the correct answer. More specifically, estimates seemed to follow logically from an erroneous interpretation of the external benchmarks provided in this condition. For example, a child estimated 103 as 153 because (s)he seems to have misallocated (the number) 100 at the 75% benchmark (i.e., the number 150). In an attempt to identify the trials on which such erroneously identified benchmarks might have been used, we calculated for each to-be-positioned number two intervals by adding and subtracting the value of the respective (misrepresented) benchmarks plus or minus ten to/from the to-be-positioned number. For instance, if a child had to estimate the position of 103, we assumed that his/her estimate would fall in the interval [43 □ 63] if (s)he erroneously represented the benchmark at 25% as 100, and in the interval [143 □ 163] if (s)he erroneously represented the benchmark at 75% as 100. Identification of these systematic errors revealed that 16% of the estimates around the first quartile, such as 43 and 55, were due to a misinterpretation of this first quartile as the midpoint (see Table 1). For the estimates around the midpoint, such as 91 and 103, results indicated that 39% of the estimates appeared to be the result of an incorrect allocation of the number 100 on the position of 150 (i.e., the third quartile). Also, 10% of the estimates around the midpoint in the quartile condition were based on the 25% benchmark. Remarkably, none of the estimates around the third quartile, such as 146 and 159, were due to incorrectly interpreting the first quartile or midpoint as the third quartile. Hence, this analysis suggested that a substantial number of children from the quartile condition associated the three given benchmarks with a wrong number to guide their estimates.

[2] We looked specifically at those three locations since they were the most likely to be wrongly interpreted due to not having the corresponding number underneath the vertical line.

(Table 1 about here)

Secondly, we looked at children's responses to the general question at the end of the interview, concerning what number was represented by the benchmarks located at 25%, 50%, and 75% of the number line (see Table 2). In the midpoint condition, 33% of the children referred to a number different from 100 when being asked which number was represented by the 50% benchmark. In the quartile condition, 86% of the children wrongly identified both the vertical line at 25% and 50%, and even 90% misinterpreted the third quartile. Interestingly, of this 90%, about half of them (i.e., 47%) identified the third quartile as representing 100 instead of 150.

(Table 2 about here)

Strategy Use

Taking into account the unexpected finding that so many children from the quartile condition were unable to identify and properly use the value of the externally provided benchmarks, we decided to exclude the data from the quartile condition from the remainder of the analyses.

Frequency of strategy use. We analyzed how often children made use of a strategy from one of the three main strategy categories in our classification scheme, and whether this varied as a function of condition. We therefore conducted a 2 (Condition: bounded vs. midpoint) × 3 (Strategy Category: perceptual strategies, SBB strategies, non-SBB strategies) ANOVA with repeated measures on the last variable and the percentage of trials being solved with a specific strategy as dependent variable. This ANOVA revealed a significant main effect of strategy category, $F(2, 80) = 32.58$, $p < .0001$. Overall, the SBB strategies (M =

55%, *SD* = 26%) were used most frequently, followed by the non-SBB strategies (*M* = 30%, *SD* = 23%) and the perceptual strategies (*M* = 12%, *SD* = 12%, all *p*s < .01).

Frequency of SBB strategies. We examined whether the use of the distinct SBB strategies differed as a function of condition and number range. Since the mean frequency of use of the 25% and the 75% SSB strategies across the bounded and midpoint condition was only 4% and 2%, respectively, we decided to exclude these strategies from the remainder of the analyses. A 2 (Condition: bounded vs. midpoint) × 3 (SBB Strategy: 0%, 50%, 100%) × 4 (Number Range: 0 □ 50, 50 □ 100, 100 □ 150, 150 □ 200) ANOVA with repeated measures on the last two variables was conducted on the frequency of strategy use. A significant main effect of strategy, $F(2, 80) = 13.17$, $p < .0001$, was found, showing that the 50% strategy (*M* = 22%, *SD* = 26%) was used significantly more than the 0% (*M* = 10%, *SD* = 20%, $p = .0001$) and 100% (*M* = 16%, *SD* = 25%, $p < .05$). We further observed a significant SBB Strategy × Condition interaction, $F(2, 80) = 10.35$, $p = .0001$, indicating that the 50% strategy was used more in the midpoint condition (*M* = 30%, *SD* = 29%) compared to the bounded condition (*M* = 15%, *SD* = 21%, $p = .001$). Also, a significant SBB Strategy × Number Range interaction was found, $F(6, 240) = 36.59$, $p < .0001$ (see Figure 4). We found that, in range 0–50, the 0% strategy (*M* = 29%, *SD* = 25%) was used significantly more frequently than the 50% (*M* = 3%, *SD* = 9%) and 100% (*M* = 2%, *SD* = 11%) strategy (both *p*s < .0001). In range 50 □ 100 and 100-150, the 50% strategy (*M* = 24%, *SD* = 22% and *M* = 45%, *SD* = 29% respectively) was used significantly more frequently than the 0% (*M* = 9%, *SD* = 17% and *M* = 3%, *SD* = 11% respectively) and 100% (*M* = 4%, *SD* = 11% and *M* = 16%, *SD* = 22% respectively) strategy (all *p*s < .05). Finally, in the range, 150 □ 200, the 100% strategy (*M* = 41%, *SD* = 30%) was used significantly more often than the 0% (*M* = 1%, *SD* = 9%) and 50% (*M* = 17%, *SD* = 22%) strategy (both *p*s < .0001). Also, the 50% strategy was used more often than the 0% strategy ($p < .05$). The three-way interaction was not significant. To summarize, the 50%

strategy was used more in the midpoint condition in comparison to the bounded condition. Moreover, children specifically relied on the origin and endpoint (and to a lesser extent also on the midpoint) in the outer ranges, whereas they especially used the 50% benchmark or midpoint in the two middle ranges.

(Figure 4 about here)

Model fittings. To test whether the use of SBB strategies was reflected in children's estimation patterns, analyses of individual estimation patterns were conducted. We separately fitted an unbounded, a one-cycle, and a two-cycle power model on the estimates of each child in the bounded and midpoint condition. The unbounded power model was also fitted to identify children who only used the origin to guide their estimates. Since the 25% and 75% strategies were used very infrequently (i.e., altogether on only 6% of all trials) and since we had left out the quartile condition in all analyses concerning strategy use, a four-cycle power model was not fitted on the data. The Akaike information criterion corrected for small samples (AICc) was used (as in Barth & Paladino, 2011; Slusser et al., 2013) to determine which model could best explain children's estimation patterns. This measure takes into account goodness of fit and model complexity (i.e., number of parameters) whereby a lower AICc value refers to a better model fit. Differences in AICc scores (i.e., ΔAICc) reflect the amount of support for one specific model in comparison to the other models. According to Burnham and Anderson (2002) models having a ΔAICc within 0-2 of the best model have substantial support and should be taken into consideration when making inferences, models with a ΔAICc within 4-7 have considerably less support and models with a ΔAICc >10 have essentially no support. Most children in the bounded condition, namely 67%,were best fit by an unbounded power model, followed by 14% and 19% for the one-cycle and two-cycle

model, respectively (see Table 3). In the midpoint condition, 48%, 14%, and 38% of the children were best fit by an unbounded, a one-cycle, and a two-cycle model, respectively. Even though we observed a decrease in the percentage of children being best fit by the unbounded power model (reflecting the strategic use of only the origin) and an increase in the percentage of children being best fit by the two-cycle model (suggesting the use of the origin, endpoint, and midpoint when making NLEs) from the bounded to the midpoint condition, a Chi-Square test failed to reveal a significant association between condition and best model fit ($\chi^2(2, N = 42) = 2.00, p > .10$).

(Table 3 about here)

Relationship between the frequency of SSB strategy use and children's NLE performance and math achievement.

To investigate the extent to which the use of SBB strategies was related to children's NLE accuracy, as well as their general mathematics achievement, we calculated the correlations between PAE, mathematics achievement, and the use of SBB strategies (see Table 4). The use of SBB strategies correlated significantly with mathematics achievement ($r = .40, p < .01$) and tended to correlate negatively with PAE ($r = -.28, p = .078$), and thus positively with the accuracy of children's NLEs.

(Table 4 about here)

Discussion

Both older and more recent sources of indirect evidence suggest that, when making NLEs, children make use of strategies based on given or self-generated benchmarks (Ashcraft

& Moore, 2012; Barth & Paladino, 2011; Ebersbach, Luwel, & Verschaffel, 2013; Newman & Berger, 1984, Petitto, 1990; Siegler & Opfer, 2003; White & Szucs, 2012). The purpose of the present study was to perform a direct and systematic analysis of children's strategy use by asking second grade children to verbally explain how they solved each NLE item, in addition to analyzing their error rates. To achieve this goal, second graders had to solve a NP-task in which the number of externally provided benchmarks was manipulated in three separate conditions: a bounded, a midpoint, and a quartile condition.

Firstly, we examined the effect of given benchmarks on children's NLE performance. The underlying idea was that, if the application of benchmark-based strategies indeed mediates children's NLE performance, then providing them with more benchmarks should lead to more frequent and more efficient strategic behavior based on these benchmarks and, consequently, to higher performance on the NLE task in terms of estimation accuracy. However, a larger number of provided benchmarks, did not have the expected positive effect on children's overall estimation accuracy. Also, we performed a contour analysis (Ashcraft & Moore, 2012), which revealed that, as hypothesized, the provision of a benchmark at the midpoint did lead to more accurate estimates for items around the middle of the number line compared to the bounded condition, but – again unexpectedly – also compared to the quartile condition. The two additional benchmarks in the quartile condition, however, did not lead to better estimates around first and third quartile in that condition. In search of an explanation for the unexpected findings for the quartile condition, we performed two additional analyses, which jointly provided strong evidence that the use of erroneously represented benchmarks led to systematic NLE errors. More specifically, children in this condition often made erroneous associations between the benchmark numbers 50, 100, and 150 on the one hand, and their corresponding location on the external number line, on the other hand.

Secondly, children's strategy use in the bounded and midpoint condition was investigated. We looked whether a larger number of provided benchmarks led to an increase in the number of strategies that were based on self-generated benchmarks or on benchmarks that were externally provided at 0%, 50%, or 100% of the number line. As expected, the 50% strategy was used more frequently in the midpoint than in the bounded condition. Furthermore, we observed the expected associations between type of SBB strategy and number range, in the sense that each type of SBB strategy was used mostly in the number range(s) where one would rationally expect it. For instance, the 0% strategy was used most frequently for estimating numbers near the origin, whereas the 50% and 100% strategy were used most frequently for estimations near the midpoint and the endpoint, respectively.

Finally, we hypothesized that the use of these benchmarks should be reflected in children's estimation patterns. By separately fitting an unbounded, a one-cycle, and a two-cycle power model on the estimates of each child in the bounded and midpoint condition, we revealed that the percentage of children being best fit by the two-cycle power model increased from the bounded (i.e., 19%) to the midpoint (i.e., 38%) condition. This result is in line with the verbal reports on children's strategy use indicating a more frequent use of the 50% strategy in the midpoint than in the bounded condition. However, we also observed a large percentage of children in the bounded and midpoint condition who were fitted best by an unbounded power model. A possible explanation for this finding might be that the unfamiliar number range used in the present study (i.e., 0-200) had prevented the strategic use of any other benchmark beyond the origin.

Thirdly, the extent to which the use of SBB strategies was related to children's NLE accuracy and their general mathematics achievement was investigated. We found a significant correlation between the use of SBB strategies and mathematics achievement and a marginally significant correlation between the use of SBB strategies and estimation accuracy. These

findings suggest that children who are more proficient in mathematics make greater use of SBB strategies for making NLEs. Moreover, the use of these SBB strategies seems to lead to more accurate estimates.

Our findings as summarized above have implications for theory, research, and educational practice.

From a theoretical perspective, our findings first of all involve a plea for taking into account more seriously participants' strategy use when solving the NLE task (Barth & Paladino, 2011; Ebersbach et al., 2013; Link et al., 2014; Slusser et al., 2013). The reported evidence that the provision of an (unlabeled) benchmark at the midpoint affects children's NLE accuracy raises doubt whether the NLE task can be considered as a pure measure of one's underlying magnitude representation. However, we also observed that additional (unlabeled) benchmarks at 25% and 75 % may have no or even a detrimental effect. A convincing explanation for this finding is still lacking, but two related post-hoc analyses suggested this might be attributed to an erroneous determination of these extra (unlabeled) benchmarks. This raises the question why children from the quartile condition departed from wrong associations between the numbers of 50, 100, and 150 on the one hand and their location on the number line on the other hand. Most probably, this wrong association might be caused by children's lacking familiarity with these numbers (Ebersbach et al., 2008), their inability to reason proportionally (Boyer, Levine, & Huttenlocher, 2008) and/or to technically execute the required multiplicative operations in the proportional relation (Barth, Baron, Spelke, & Carey, 2009). For instance, Boyer et al. (2008) showed that 10- to 12-year olds have difficulty solving proportional reasoning problems when the proportions are represented in discrete quantities. Furthermore, it could be argued that children's familiarity with the base 10 system would lead to better estimations with benchmarks representing decades (i.e., 10, 20, …) rather than larger quantities such as multiples of 50. A possible way to further

investigate this issue would be to compare children's NLE performance and strategy use when confronted with a bounded number line, a number line with a benchmark at each decile, and a number line with benchmarks at each quartile.

Interestingly, Siegler and Thompson (2014) found that not all types of benchmarks had a beneficial effect on fifth graders' performance when estimating the position of common fractions on a number line. More specifically, external benchmarks that divided a 0 □ 1 number line into tenths, quarters, or fifths led to less accurate estimates than having no benchmark or a midpoint benchmark. It was demonstrated that the decile, quartile, and quintile benchmarks led to an improper encoding of the fractions on the basis of its numerator or denominator rather than on the fraction magnitude, which in its turn led to less accurate estimates. Obviously, this explanation cannot account for the present findings since all the presented numbers in our study were whole numbers and thus children could not encode the presented numbers in terms of numerator and denominator but only as a magnitude.

From a methodological perspective, we point out that we tried investigating children's strategies for NLE by collecting verbal reports. The actual distribution of the different SBB strategies across the different ranges of the number line provides additional support for the validity of these reports. Nevertheless, given that these verbal methods have their well-known limitations too (Crutcher, 1994; Ericsson & Simon, 1980), we suggest to collect other kinds of data on NLE strategies such as (video-based) observations of children's actual solution behavior, their pointing behavior on the number line (for instance, when presented on a tablet; Vermeulen, Scheltens, & Eggen, submitted), and/or their eye-movements (Schneider et al., 2009).

We finally turn to some implications for early and elementary mathematics education. First, our research suggests that it is possible to improve the accuracy of children's NLEs by providing them external benchmarks, although our study also yielded strong evidence that this

instructional intervention may also lead to unexpectedly negative results, particularly for benchmarks that go beyond the 50% benchmark and when working in a number domain with which the learners are not yet completely familiar. Anyhow, when adding such extra benchmarks, it is recommendable to check carefully if learners interpret and represent them properly. More generally, our study suggests that it would be interesting to design and evaluate instructional environments that stimulate the development of children's NLEs. Arguably, several such environments have already been developed, typically in a game-based environment (Ramani, Siegler, & Hitti, 2012; Whyte & Bull, 2008) and some of these studies have already yielded promising results. However, to the best of our knowledge, so far no such intervention has paid pivotal attention at the development of children's estimation strategies by working intentionally and systematically at children's underlying (benchmark-based) strategies. However, it should be noted that, given the results of the present study, these interventions might improve children's NLE performance but this improvement does not necessarily entail an improvement in children's underlying magnitude representation.

References

Ancker, J.S., & Kaufman, D. (2007). Rethinking health numeracy: A multidisciplinary literature review. *Journal of the American Medical Informatics Association, 14,* 713-721.

Ashcraft, M.H., & Moore, A.M. (2012). Cognitive processes of numerical estimation in children. *Journal of Experimental Child Psychology, 111,* 246-267.

Barth, H.C., Baron A., Spelke E., & Carey S. (2009). Children's multiplicative transformations of discrete and continuous quantities. *Journal of Experimental Child Psychology, 103,* 441–454.

Barth, H.C., & Paladino, A.M. (2011). The development of numerical estimation: Evidence against a representational shift. *Developmental Science*, *14*, 125-135.

Berteletti, I., Lucangeli, D., Piazza, M., Dehaene, S., & Zorzi, M. (2010). Numerical estimation in preschoolers. *Developmental Psychology, 46,* 545-551.

Booth, J.L., & Siegler, R.S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology*, *41,* 189-201.

Booth, J.L., & Siegler, R.S. (2008). Numerical magnitude representations influence arithmetic learning. *Child Development, 79,* 1016–1031.

Boyer T.W., Levine S.C., & Huttenlocher J. (2008). Development of proportional reasoning: Where young children go wrong. *Developmental Psychology, 44,* 1478-1490.

Bugden, S., & Ansari, D. (2011). Individual differences in children's mathematical competence are related to the intentional but not automatic processing of Arabic numerals. *Cognition, 118,* 32-44.

Burnham, K.P., & Anderson, D.R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd edn.). New York: Springer.

Crutcher, R.J. (1994). Telling what we know: The use of verbal report methodologies in psychological research. *Psychological Science, 5,* 241-245.

De Smedt, B., Verschaffel, L., & Ghesquière, P. (2009). The predictive value of numerical magnitude comparison for individual differences in mathematics achievement. *Journal of Experimental Child Psychology, 103,* 469-479.

Dehaene, S. (1997). *The number sense.* Oxford: Oxford University Press.

Dudal, P. (2000). *Leerlingvolgsysteem: Wiskunde — Toetsen 1-2-3. Basisboek [Student monitoring system: Mathematics — Tests 1-2-3 manual].* Leuven, Belgium: Garant.

Ebersbach, M., Luwel, K., Frick, A., Onghena, P., & Verschaffel, L. (2008). The relationship between the shape of the mental number line and familiarity with numbers in 5- to 9-year old children: Evidence for a segmented linear model. *Journal of Experimental Child Psychology, 99,* 1-17.

Ebersbach, M., Luwel, K., & Verschaffel, L. (2013). Comparing apples and pears in studies on magnitude estimates. *Frontiers in Cognitive Science, 4,*1-6.

Ericsson, K.A., & Simon, H.A. (1980). Verbal reports as data. *Psychological Review, 87,* 215-251.

Ericsson, K.A., & Simon, H.A. (1984). *Protocol analysis: Verbal reports as data.* Cambridge, MA: Bradford Books/MIT Press.

Finnie, R., & Meng, R. (2001). Cognitive skills and the youth labour market. *Applied Economics Letters, 8*, 675-679.

Gandini, D., Ardiale, E., & Lemaire, P. (2010). Children' strategies in approximate quantification. *Current Psychology Letters: Behaviour, Brain, & Cognition, 26,* 1-14.

Gandini, D., Lemaire, P., & Dufau, S. (2008). Older and young adults' strategies in approximative quantification. *Acta Psychologica, 129,* 175-189.

Gilmore, C.K., McCarthy, S.E., & Spelke, E.S. (2010). Non-symbolic arithmetic abilities and mathematics achievement in the first year of formal schooling. *Cognition, 115,* 394-406.

Landis, J.R., & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33,* 159-174.

Link, T., Huber, S., Nuerk, H.-C., & Moeller, K. (2014). Unbounding the mental number line - new evidence on children's spatial representation of numbers. *Frontiers in Psychology, 4*, 1-12.

McGilly, K., & Siegler, R.S. (1989). How children choose among serial recall strategies. *Child Development, 60,* 171-182.

Moeller, K., Pixner, S., Kaufmann, L., & Nuerk, H.-C. (2009). Children's early mental number line: Logarithmic or decomposed linear? *Journal of Experimental Child Psychology, 103*, 503-515.

Newman, R.S., & Berger, C.F. (1984). Children's numerical estimation: Flexibility in the use of counting. *Journal of Educational Psychology, 76,* 55-64.

Petitto, A.L. (1990). Development of number line and measurement concepts. *Cognition and Instruction, 7,* 55-78.

Ramani, G.B., Siegler, R.S., & Hitti, A. (2012). Taking it to the classroom: Number board games as a small group learning activity. *Journal of Educational Psychology, 104,* 661-672.

Sasanguie, D., De Smedt, B., Defever, E., & Reynvoet, B. (2012). Association between basic numerical abilities and mathematics achievement. *British Journal of Developmental Psychology, 30*, 344-357.

Schneider, M., Grabner, R.H. & Paetsch, J. (2009). Mental number line, number line estimation, and mathematical school achievement: Their interrelations in Grades 5 and 6. *Journal of Educational Psychology, 101,* 359-372.

Siegler, R.S., & Booth, J.L. (2004). Development of numerical estimation in young children. *Child Development, 75,* 428–444.

Siegler, R.S., & Opfer, J.E. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science, 14,* 237-243.

Siegler, R.S., & Stern, E. (1998). Conscious and unconscious strategy discoveries: A microgenetic analysis. *Journal of Experimental Psychology: General, 127,* 377–397.

Siegler, R.S., Thompson, C.A., & Opfer, J.E. (2009). The logarithmic-to-linear shift: One learning sequence, many tasks, many time scales. *Mind, Brain, and Education, 3,* 143-150.

Siegler, R.S., & Thompson, C.A. (2014). Numerical landmarks are useful – except when they're not. *Journal of Experimental Child Psychology, 120,* 39-58.

Slusser, E.B., Santiago, R.T., & Barth, H.C. (2013). Developmental change in numerical estimation. *Journal of Experimental Psychology: General, 142*, 193-208.

Thompson, C.A., & Opfer, J.E. (2010). How 15 hundred is like 15 cherries: Effect of progressive alignment on representational changes in numerical cognition. *Child Development, 81,* 1768-1786.

Torbeyns, J., De Smedt, B., Ghesquière, P., & Verschaffel, L. (2009). Acquisition and use of shortcut strategies by traditionally schooled children. *Educational Studies in Mathematics, 71,* 1-17.

Vermeulen, J.A., Scheltens, F., Eggen, T.J.H.M., (2014). Strategie-identificatie met de lege getallenlijn: Een vergelijking tussen tablets en papier [Strategy identification on the

empty number line: A comparison between tablets and paper]. Manuscript submitted for publication.

White, S.L.J., & Szucs, D. (2012). Representational change and strategy use in children's number line estimation during the first years of primary school. *Behavioral and Brain Functions, 8,* 1-12.

Whyte, J.C., & Bull, R. (2008). Number games, magnitude representation, and basic number skills in preschoolers. *Developmental Psychology, 44*, 588–596.

Figure 1. Presented number line in (a) bounded, (b) midpoint, and (c) quartile condition.

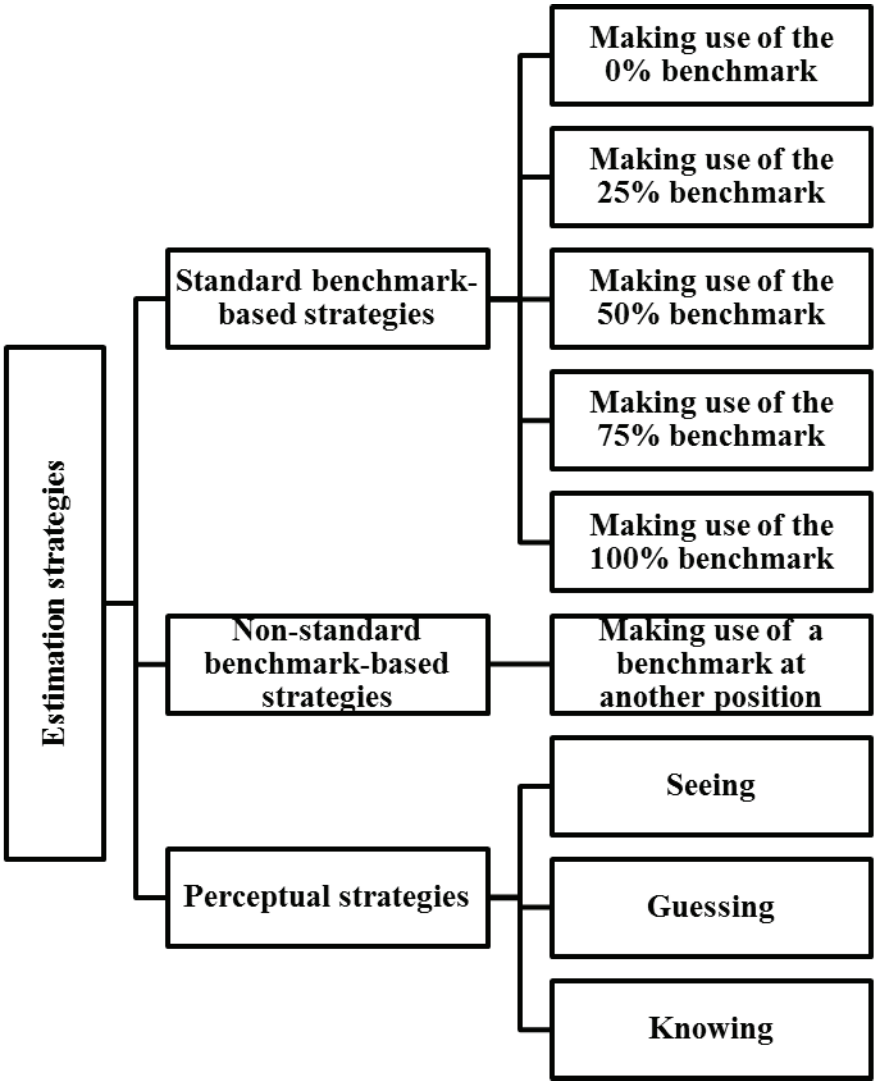

Figure 2. Decision tree for the classification of NLE strategies.

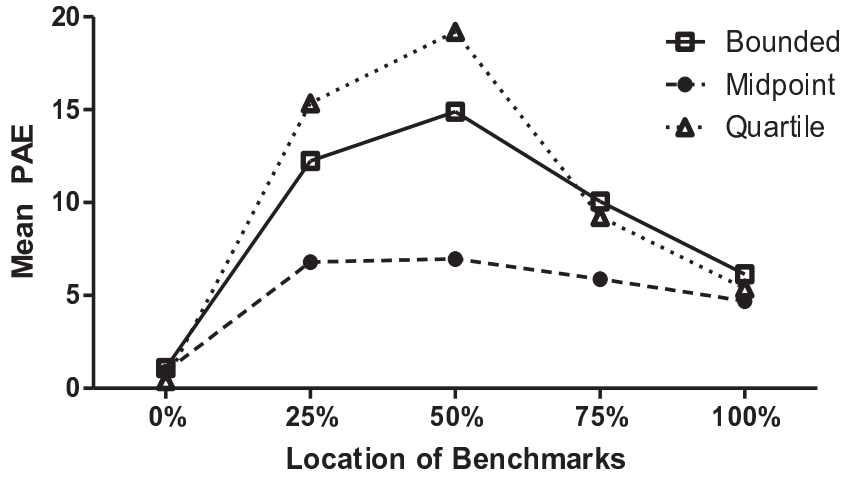

Figure 3. Mean percentage of absolute error (PAE) around the standard benchmarks as a function of condition.

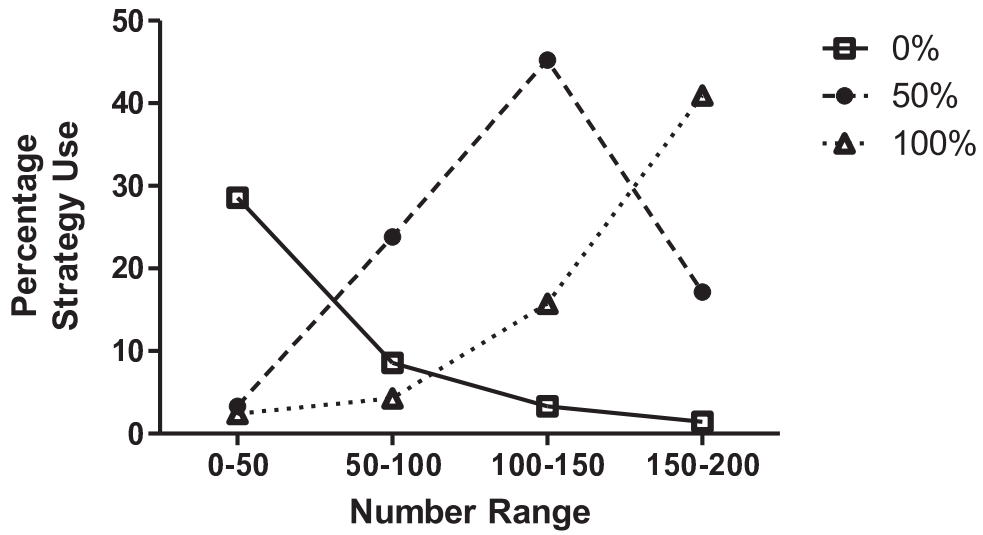

Figure 4. Percentage use of the 0%, 50%, and 100% SBB strategies for each range.

Table 1

Percentage of Estimates Around the Three Benchmarks Based on Wrongly-used Benchmarks for the Quartile Condition.

Benchmark at	Wrongly interpreted as		
	25%	50%	75%
25%		16% (n = 7)	2% (n = 1)
50%	10% (n = 4)		39% (n = 16)
75%	0%	0%	

Note. For instance, 39% of the trials surrounding the 50% benchmark (i.e., 100) were estimated around the 75% benchmark due to wrongly locating the number 100 on the position of 150 (i.e., 75% benchmark).

Table 2

Percentage of Wrongly-identified Benchmarks for the Midpoint and Quartile Condition based on Interview Data

	Benchmarks at		
Condition	25%	50%	75%
Midpoint		33% (n = 7)	
Quartile	86% (n = 18)	86% (n = 18)	91% (n = 19)

.

Table 3

Percentage of Children Best Fit by Each Power Model and the Difference Scores for AICc in the Bounded and Midpoint Condition.

	Bounded condition		Midpoint condition	
Power models	%	ΔAICc	%	ΔAICc
Unbounded	67 (n = 14)	one-cycle: 11.96 two-cycle: 15.87	48 (n = 10)	one-cycle: 10.58 two-cycle: 13.00
One-cycle	14 (n = 3)	unbounded: 3.56 two-cycle: 9.63	14 (n = 3)	unbounded: 2.96 two-cycle: 2.19
Two-cycle	19 (n = 4)	unbounded: 5.95 one-cycle: 5.00	38 (n = 8)	unbounded: 9.70 one-cycle: 9.48

Table 4

Correlations between NLE Accuracy, Mathematics Achievement, and SSB Strategy Use

	1.	2.
1. NLE accuracy (PAE)		
2. Mathematics achievement	-.12	
3. % Total SBB strategy use	-.28	.40*

Note. *$p < .01$.